


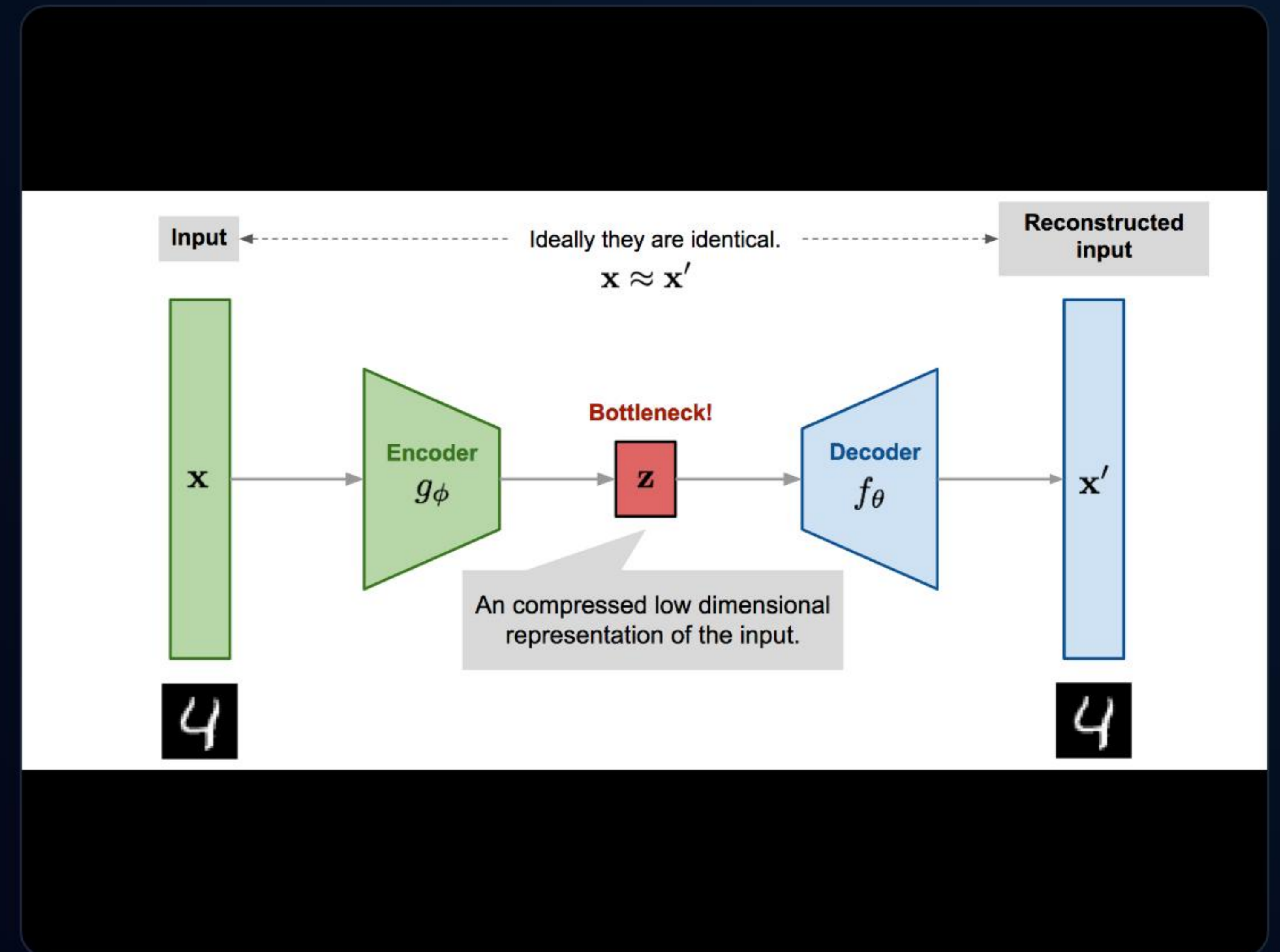

The VAE Pipeline

Encoding Data as Probability Distributions

System Architecture

-  **Encoder:** Parametrizes the variational distribution $q_{\phi}(\mathbf{z} | \mathbf{x})$ to infer latent code variables.
-  **The Bottleneck:** Samples a vector \mathbf{z} from the inferred distribution, representing input \mathbf{x} as a range.
-  **Decoder:** Maps latent samples \mathbf{z} back to the input space using the generative model $p_{\theta}(\mathbf{x} | \mathbf{z})$.



The Gaussian Prior Constraint

Global Anchor





We force the latent space to follow a **Standard Normal Distribution**. This prevents the encoder from scattering data arbitrarily.

$$p(z) \sim N(0, I)$$

Clustering Effect

The prior pulls latent codes toward a shared, **zero-centered** region. This ensures that every sample resides in a high-density, generative manifold.

| What is the ELBO Objective?

-  **Evidence Lower Bound:** A surrogate function used because calculating the actual data log-likelihood $\log p(\mathbf{x})$ is mathematically intractable.
-  **Safe Proxy:** ELBO provides a lower bound for the likelihood; maximizing it indirectly optimizes the model's predictive power.
-  **Joint Optimization:** It enables training both the encoder and decoder simultaneously using stochastic gradient descent.
-  **The Balance:** It consists of the Reconstruction Log-Likelihood and the **KL Loss** penalty.

The ELBO Loss Function

$$\text{ELBO} = \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x|z)] - \underbrace{D_{\text{KL}}(q_{\phi}(z|x) \parallel p(z))}_{\text{KL Loss}}$$

Reconstruction Term

Forces the decoder to recreate the input with high fidelity. (e.g., MSE or BCE Loss).

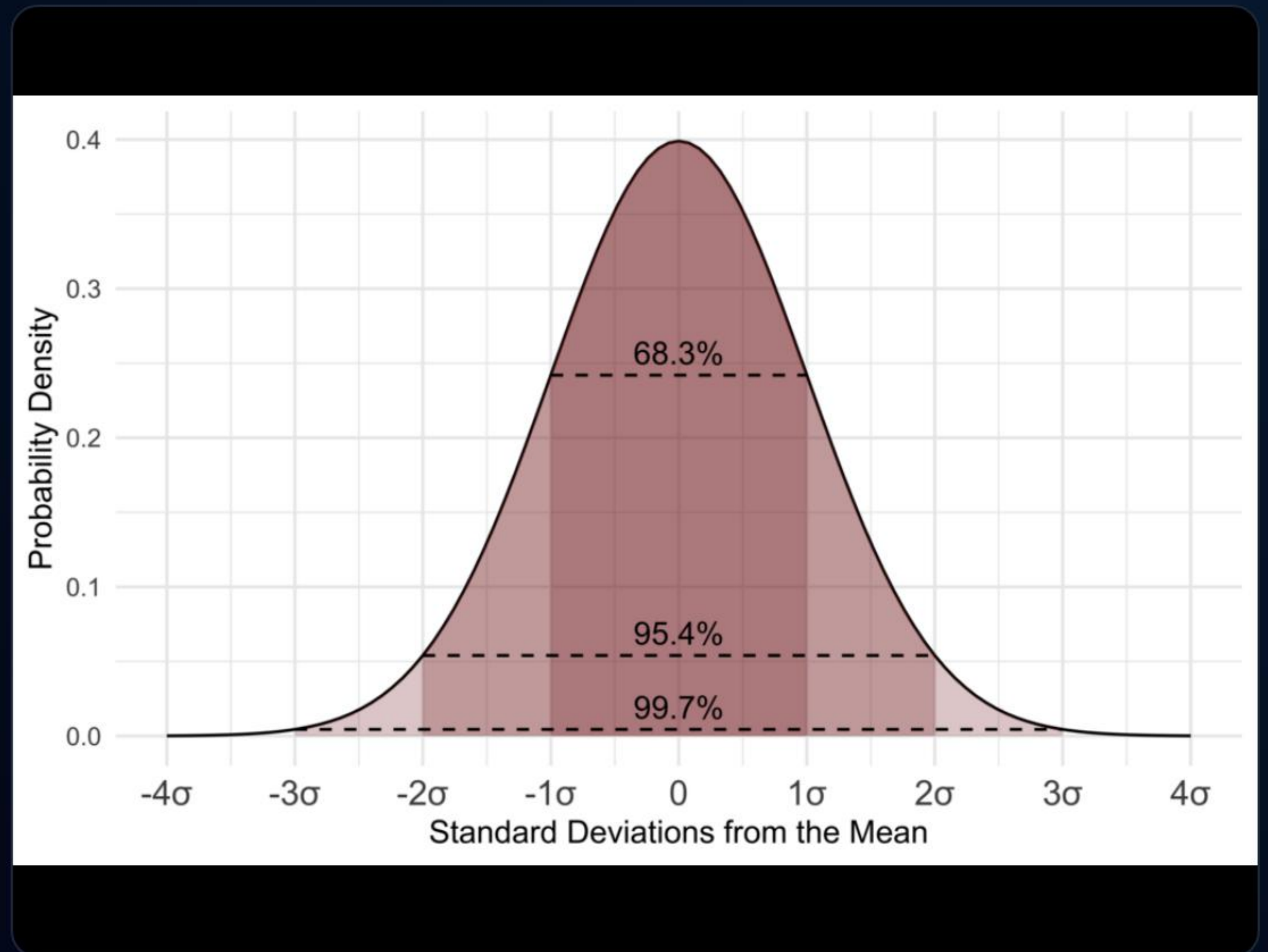
KL Loss Term

Regularizes the encoder to stay close to the simple Gaussian prior distribution.

The Role of KL Loss

The **KL Loss** acts as a regularizing force that prevents the encoder from overfitting to individual points.

- ⚡ **Centering:** Pulls latent means μ toward 0.
- 📏 **Scaling:** Forces variance σ^2 toward 1.
- 🔗 **Overlap:** Encourages distributions to touch, ensuring continuity.



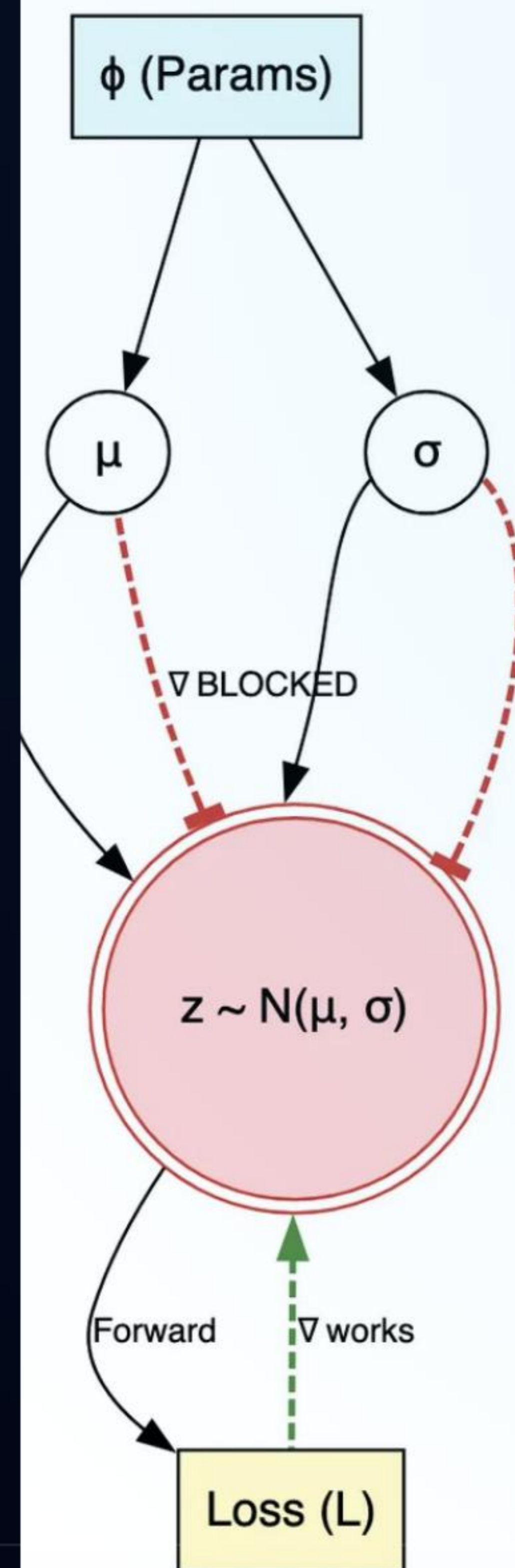
Reparameterization

Sampling is non-differentiable, creating a "break" in backpropagation. To fix this, we move randomness outside the path.

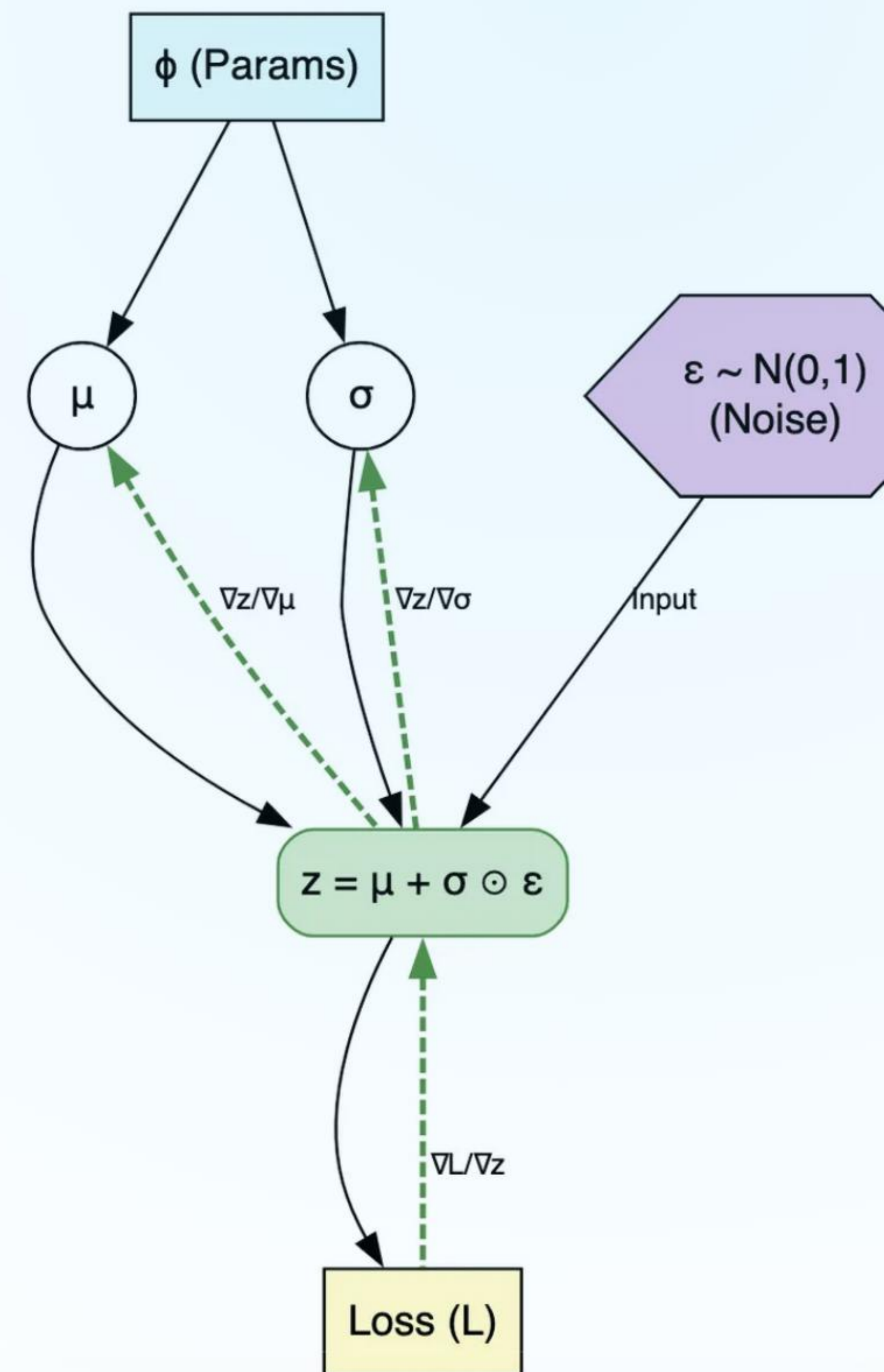
$$z = \mu + \sigma \odot \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I})$$

The network now only learns the parameters μ and σ deterministically, while ε provides the noise.

A. Standard Stochastic Node
(Gradient Blocked)



B. Reparameterization Trick
(Gradient Flows)

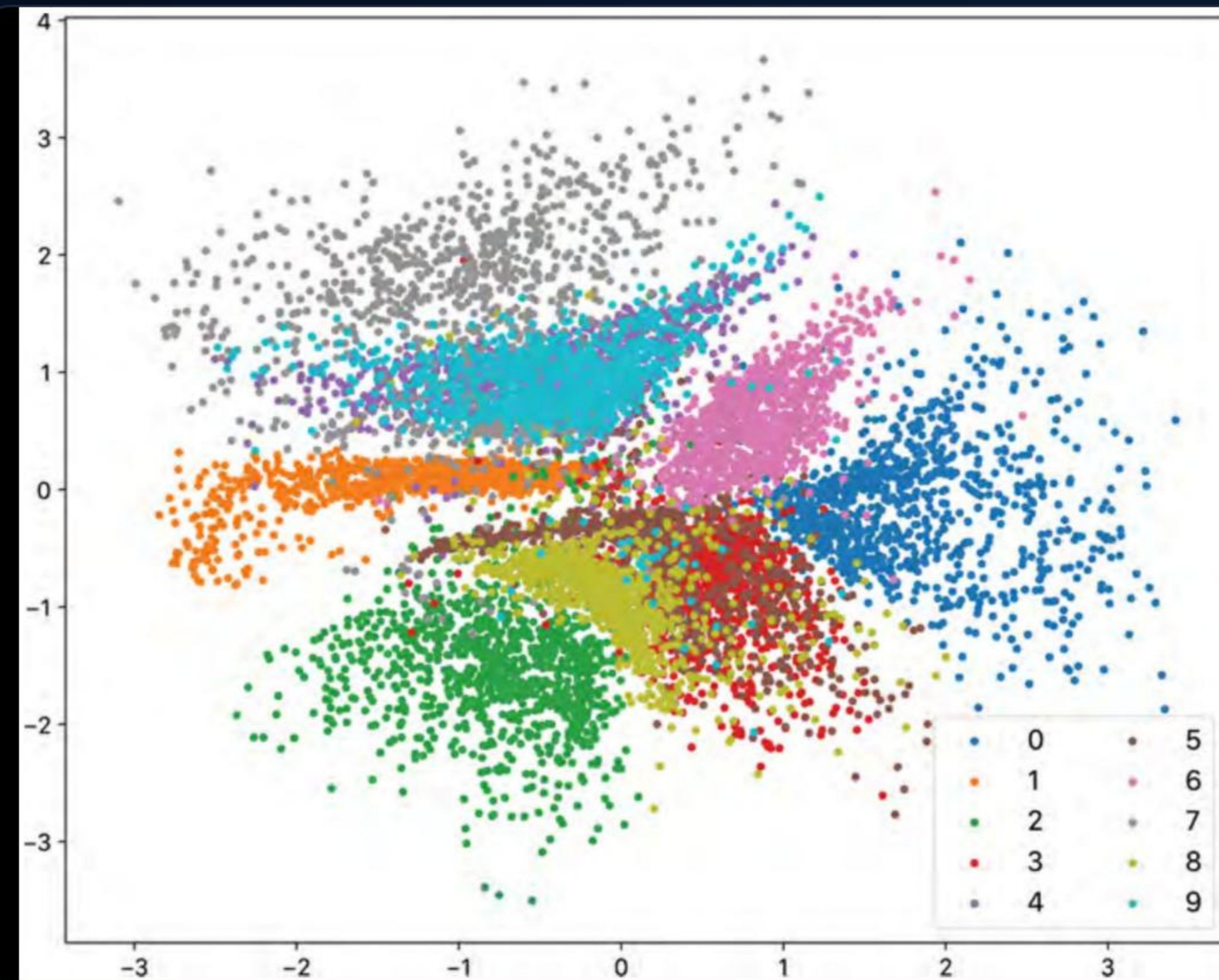


Latent Manifold Continuity

Smooth Transitions

Because the **KL Loss** forces distributions to overlap, moving through latent space results in smooth transformations.

Interpolation between points reveals how the model has learned the underlying semantic structure of the data.



Optimization Trade-offs

Weighting	Latent Outcome	Output Quality
High Likelihood	Isolated clusters; sparse space.	Sharp reconstructions, poor generation.
High KL Loss	Dense, centered Gaussian cloud.	Blurry results (Posterior Collapse).
Balanced ELBO	Structured & continuous manifold.	Sharp, diverse, and sampleable.

Optimal training requires careful negotiation between these two competing terms.

| Summary Comparison

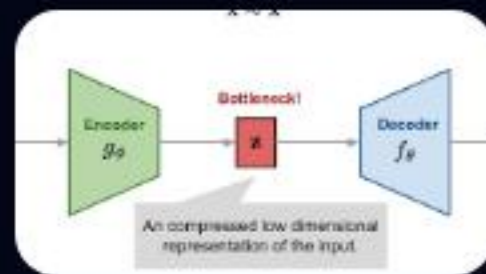
Standard Autoencoder

Deterministic bottleneck mapping input to a single point. Non-probabilistic space leads to empty gaps where generation is impossible.

Variational VAE

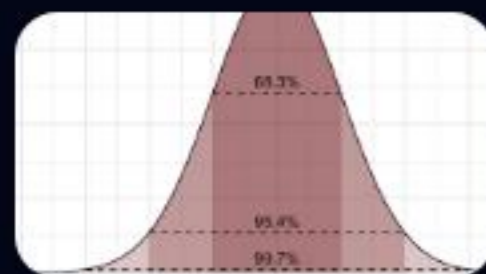
Probabilistic bottleneck mapping input to a distribution. KL-anchored manifold enables smooth sampling for generative tasks.

Image Sources



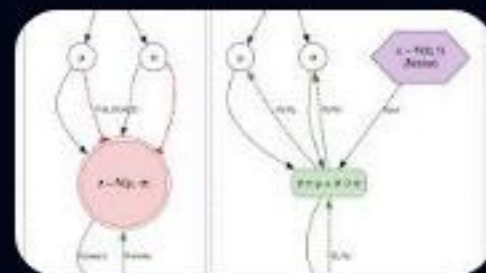
<https://lilianweng.github.io/posts/2018-08-12-vae/autoencoder-architecture.png>

Source: lilianweng.github.io



https://www.statisticshowto.com/wp-content/uploads/2024/06/Standard_Normal_Distribution.png

Source: www.statisticshowto.com



<https://hunterheidenreich.com/img/notes/variational-autoencoder-reparameterization-trick.webp>

Source: hunterheidenreich.com



https://datahacker.rs/wp-content/uploads/2026/03/fig_08_vae_2d_latent_space-scaled.jpg

Source: datahacker.rs
